



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

Address: COMMISSIONER FOR PATENTS

P.O. Box 1450

Alexandria, Virginia 22313-1450

www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|-----------------|-------------|----------------------|---------------------|------------------|
|-----------------|-------------|----------------------|---------------------|------------------|

10/631,172

07/31/2003

James Michael McArdle

AUS920030382US1

1419

45993

7590

10/07/2008

IBM CORPORATION (RHF)

C/O ROBERT H. FRANTZ

P. O. BOX 23324

OKLAHOMA CITY, OK 73123

EXAMINER

HOFFLER, RAHEEM

ART UNIT

PAPER NUMBER

2165

MAIL DATE

DELIVERY MODE

10/07/2008

PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary

Application No.

10/631,172

Applicant(s)

MCARDLE, JAMES MICHAEL

Examiner

RAHEEM HOFFLER

Art Unit

2165

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 12 August 2008.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-5 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-5 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☐ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-946)
- 3) ☐ Information Disclosure Statement(s) (PTO/SF/ICE)
Paper No(s)/Mail Date _____
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date _____
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: _____

Detailed Action

Continued Examination Under 37 CFR 1.114

A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 12 August 2008 has been entered. Claims 6-18 have been cancelled. Due to the amendment, Examiner withdraws rejection of claims 10 & 15 under 35 USC 112 2nd paragraph, as well as rejection under 35 USC 101.

Claim Rejections - 35 USC § 112

The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

Claim1 is rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. The scope of "generating a set of cleaning attributes for each cleaned data record in a complete set of data records...modified by a previous cleaning operation on a set of data records" is not defined within the Specification. The "previous cleaning operation" is unknown, at least in regards to Applicant's claim language and it

would be impossible to determine what the cleaning attributes are because of their dependence on the "previous cleaning operation".

The third and fourth limitations of Claim 1, "determining a degree of correlation..." and "responsive to said degree of correlation exceeding a threshold...", are also hindered by the object of "said previous cleaning operation" because their outcomes are indefinitely dependent on the object "previous cleaning operation" such as "declaring said data feature...suspect".

The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claim 1 is rejected under 35 U.S.C. 112, first and second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

The phrase "said cleaning attributes reflecting which fields of each record have been modified by a previous cleaning operation on a set of data records" is indefinite because it is unclear to Examiner as to what the "previous cleaning operation" is and how "cleaning attributes" interact with "a previous cleaning operation".

The newly added limitation, "said data feature appearing in said previously-cleaned data records" is unclear. It is not clearly defined within the claim language as to what element is referenced by the amended language.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 1-5 are rejected under 35 U.S.C. 103(a) as being unpatentable over Burdick et al (USPG Pub No. 20040107203A1; Burdick hereinafter) in view of Applicant admitted prior art (Background: Figure 1 & 2; see paragraphs [0001-0031]; Background hereinafter)

As for Claim 1, Burdick et al teaches,

“generating a set of cleaning attributes for each cleaned data record in a complete set of cleaned data records, said cleaning attributes reflecting which fields of each record have been modified by a previous cleaning operation on a set of data records” (see Fig. 1; see paragraph [0038]; e.g., attribute of an entity), and

“determining a degree of correlation of said data feature to fields of said subset of cleaned data records reflected by said cleaning attributes” (see paragraph [0035]; e.g., determining if two records are duplicates involves performing a similarity test that qualifies the similarity of two records),

“responsive to said degree of correlation exceeding a threshold, declaring said data feature appearing in said previously-cleaned data records as suspect due to said previous cleaning operation and as having been modified by said previous cleaning

operation" (Examiner refuses to examine said limitation because it cannot be implemented according to a lack of support within Applicant's specification, as necessitated by rejection under 35 USC 112, 1st and 2nd paragraph).

Burdick fails to explicitly recite, "...receiving a data feature identified within said cleaned data record by a data mining process for a subset..."

Applicant's background explicitly recites, "receiving a data feature identified within said cleaned data record by a data mining process for a subset of said complete set of cleaned data records" (see paragraph [0010-0017; e.g., utilizing data mining tools and techniques to return particular information from said complete set of cleaned data).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to have combined the method and system of data analysis taught within Applicant's background with the data cleansing application of Burdick in order to produce much more accurate and efficient results than could be obtained previously (see Burdick [0008]).

As for Claim 2, Burdick teaches, forming "...a cleaning attributes register for each cleaned data record" (see paragraph [0057-0058]), and "generating a set of bit-mapped Boolean flags" (see paragraph [0058]; e.g., Boolean expressions utilized for the data cleaning process.)

As for Claim 3, Burdick teaches, "performing an operation selected from the group of appending a set of cleaning attributes to each cleaned data record, prepending

a set of cleaning attributes to each cleaned data record, distributing a set of cleaning attributes to each cleaned data record, and generating a cleaning attribute table" (see Figure 1; see paragraph [0034], [0038]; e.g., attributes of an entity).

As for Claim 4, Burdick teaches, "a step selected from a group of receiving a cluster, receiving a trend, and receiving a pattern" (see paragraph [0026], [0032]; e.g., clustering, matching, standardization).

As for Claim 5, Burdick teaches, "comparing each record in a raw data set to each record in a cleaned data set" (see paragraph [0069-0070]).

Response to Arguments

Applicant's arguments with respect to claims 1-5 have been carefully considered but are not persuasive in view of the original grounds of rejection.

With respect to Applicant's argument that:

"We respectfully disagree that Burdick teaches declaring "highly modified data from a cleaning process as being suspect" at their paragraphs [0035], [0053]...we believe Burdick is merely detecting duplicates in the *uncleaned* data in a step to "pre-process" the data *before* cleaning is to be performed (para. 0051). Thus, their process would be unable to determine or declare a post-cleaning data feature as "suspect" as a result of previous cleaning operations.

We respectfully submit that Burdick's steps #302 - #305 (para. 0051) are part of a *preprocessing component* #202 (para. 0048), which are not performed *after* cleaning the data, but instead are performed *before* cleaning the data in order to prepare to perform cleaning, hence, their term pre-processing. Thus, the data could not be declared as "suspect" *as a result of cleaning* because cleaning has not been performed yet."

Examiner is not persuaded. As stated within paragraph [0026], there are a number of different processes that come along with cleansing data such as parsing, validation/correction, standardization and clustering. These steps aid in accumulating a similarity score which determines which records stand out, or are deemed suspect. As stated within paragraph [0035] and further elaborated within [0036], records with a similarity score greater than a certain threshold are considered "duplicates", and therefore "suspect" by the standards of this reference. Paragraph [0053] merely teaches of a process of ensuring the quality of output for records placed within the series of steps discussed earlier.

With respect to Applicant's argument that:

"Regarding Burdick's Figure 1, the term 'attribute of an entity' does not appear, nor does 'cleaning attribute' or just 'clean'. In fact, as best we can tell, the only 'attribute' mentioned in Figure 1 is the 'violated attribute dependencies' (eleventh row). We ask the Examiner to reconsider whether or not Figure 1 is actually illustrating a database (e.g. rows being records, columns being fields), or whether it is a tabular illustration of the *kinds* of problems that can occur in data and their possible solutions. Burdick's paragraph 0003 states only that Figure 1 illustrates the different 'factors' which may result in dirty data, but it is silent regarding that Figure 1 illustrates actual dirty data (e.g. actual database contents)"

Examiner is not persuaded. Figure 1 is an illustration providing information about the various records being placed through a series of steps in order to "clean" data that is considered "dirty". Figure 1 illustrates examples of "attributes" that are analyzed in order to reach this goal by demonstrating what can qualify as being "dirty" within a database record.

With respect to Applicant's argument that:

"Without our cleaning attributes which indicate *which fields* in each record have been modified by previous cleaning operations, Burdick cannot possibly or logically teach any other operations which depend on those cleaning attributes. So, for at least the foregoing reasons, we respectfully disagree with the Examiner's position that Burdick teaches this step at their para. 0035:

We believe this is merely a duplication detection method they are disclosing, but they are not correlating that detected duplication to whether or not those fields were changed by previous data cleaning. To know this, one would need an indicator of whether not those duplicate records had been modified by the previous cleaning operation, i.e. our cleaning attributes.

But, they have no field cleaning attributes, and thus their algorithm will simply delete the "duplicates" as they have determined them to be."

Examiner is not persuaded. Figure 1 is a clear indication that detecting duplicate records is not the primary objective of the Burdick reference. Since Figure 1 provides examples of what type of value errors are being sought for correction and cleansing, it should be clearly seen by Applicant that calculating a degree of similarity between records is just one key method of Burdick.

Conclusion

The prior art made of reference and not relied upon is considered pertinent to Applicant's disclosure.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to RAHEEM HOFFLER whose telephone number is (571)270-1036. The examiner can normally be reached on 7:30 a.m. - 5:00 p.m..

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Christian Chace can be reached on (571) 272-4190. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/R. H./
Examiner, Art Unit 2165
/H. Q. P./
Primary Examiner, Art Unit 2169
/Christian P. Chace/

Supervisory Patent Examiner, Art Unit 2165

Application Number**Application/Control No.**

10/631,172

Examiner

RAHEEM HOFFLER

**Applicant(s)/Patent under
Reexamination**

MCARDLE, JAMES MICHAEL

Art Unit

2165